Statistical Hypothesis Testing and Private Information*

Aleksey Tetenov†

Jan 1, 2014

Abstract

This paper provides a rationalization for using classical one-sided hypothesis tests for approving innovations, such as new drugs or treatment programs. I consider statistical testing as a strategy of the regulator in a game against informed proponents who can profit from the regulator's acceptance decision even for ineffective innovations. Lower test size deters proponents from attempting costly clinical trials to gain approval of innovations they know ex ante to be ineffective. I show that test size equal to the ratio of the proponents' cost of collecting data to their benefit from approval of a null innovation achieves minimax optimality for a regulator who prefers not to place any prior on the quality of potential innovations. It is also the limit of decision rules for Bayesian regulators as the assumed share of potential proponents with bad innovations converges to one.

1 Introduction

The practice of statistical hypothesis testing has been widely criticized for decades across many fields that use it. Examples of such criticism include Cohen (1994), Johnson (1999), and Ziliak and McCloskey (2008). While conventional test levels of 1%, 5% and 10% are widely agreed upon, these numbers lack any substantive motivation, yet their arbitrary choice affects thousands of influential decisions. This standard method also has an unusual lexicographic structure: first, ensure that the probability of Type I errors does not exceed a given conventional level under the null hypothesis, then do the best you can to reduce the probability of Type II errors. It is difficult to motivate this structure by considering statistical decision problems. Instead,

*Thanks to Larbi Alaoui, V. Bhaskar, Ivan Canay, Toomas Hinnosaar, Chuck Manski, Ignacio Monzon, Denis Nekipelov, Amedeo Piolatto and Adam Rosen for helpful comments and suggestions. I've benefited from presenting this paper at the Cornell Conference on Partial Identification and Statistical Decisions, Collegio Carlo Alberto, Oxford, UCL, and at ASSET 2013. Part of this research was conducted at UCL/CeMMAP with financial support from the European Research Council grant ERC-2009-StG-240910-ROMETA.

†Collegio Carlo Alberto, Moncalieri (Torino), Italy. aleksey.tetenov@carloalberto.org

both Bayesian and frequentist criteria generally seek to balance in some way the probabilities of both Type I and Type II errors.

Accepting or rejecting innovations is one of the contexts in which hypothesis testing is widely used. It is prescribed, for example, by international guidelines for drug approval (International Committee on Harmonisation, 1998). Hypothesis testing is also implicitly used when evidence supporting a proposed policy is brushed away for not being "statistically significant." When a choice between two options has to be made at a given time, Simon (1945) argued that it would be more sensible simply to choose the one favored by the available evidence, even if only by a small margin. This decision rule is essentially a one-sided test with 50% level. A recent line of literature applying the minimax regret criterion to treatment choice problems reaches similar conclusions (Manski, 2004; Manski and Tetenov, 2007; Schlag, 2007; Hirano and Porter, 2009; Stoye, 2009). As decision rules, standard hypothesis tests appear too biased in favor of the status quo. Loss aversion could motivate decision rules similar to hypothesis tests, but tests with conventional levels seem too biased to be rationalized by loss aversion (Tetenov, 2012).

In this paper, I propose to study hypothesis tests as a strategy in a game against economically motivated proponents, rather than in a game against nature. Decision rules that treat the status quo and the innovation symmetrically could be problematic for the regulator by adversely affecting the mix of innovations that are proposed and tested. This important consideration is absent in analysis of games against nature. In many applications of statistical testing, proponents of innovations have private information about their quality that is not accessible to the regulator. Proponents often stand to benefit from the approval decision even if the innovation is bad from the regulator's perspective. If bad innovations faced a higher probability of acceptance based on the regulator's statistical tests, more of them would be proposed.

While the supply of positive innovations is limited, it may be reasonable to consider the supply of bad innovations almost unbounded. Anyone could come up, for example, with thousands of "ideas" of educational reforms aimed at improving student performance. The "innovator" may know ex ante that these ideas have no merit, but reliance on statistical evidence means that with some probability the evidence turns out to be strong enough for them to be accepted by the regulator. The regulator could face steep losses if her statistical tests do not deter such opportunistic "innovators" from trying their luck.

I restrict attention to the problem of statistical testing of innovations, in which the only tool available to the regulator is to accept or reject them based on credible data. To deter the proponents of bad ideas, the probability of Type I errors (acceptance of bad innovations) has to be smaller than the ratio between the proponent's cost of collecting the evidence (e.g., clinical trials or acquiring other data that may provide supporting evidence for the idea) and the proponent's benefit from having the bad innovation accepted. This framework provides a decision-theoretic rationale for using one-sided hypothesis testing in some statistical treatment choice problems and a rule for picking the test level.

Allowing for a bottomless supply of bad ideas may seem far-fetched. I show that the same solution could be obtained as a limit case if the regulator places a prior distribution over the quality of potentially proposed innovations. As the probability assigned to bad innovators increases, the regulator's optimal policy converges to the one-sided hypothesis test rule.

A couple of recent papers consider related problems of regulatory approval. Manski (2013) also advocates using randomized regulatory approval for deterring unwanted applications (as well as for diversification and learning). In contrast to this paper, randomization proposed Manski (2013) depends on readily observable characteristics of applications, rather than on statistical signals, and does not seek to reveal applicants' private information. Ottaviani and Henri (2013) study the problem of sequential research investment by an agent (e.g., a pharmaceutical company) indirectly controlled by regulatory approval. In their setting, the agent and the regulator share common beliefs about the probability of two possible states of the world.

The paper proceeds as follows: in the next section, I discuss why hypothesis testing criterion has been difficult to rationalize in statistical decision theory. Section 3 outlines a simple environment that motivates hypothesis testing rules as a minimax strategy for the regulator. Section 4 considers the problem with an additional assumption that the data satisfies the monotone likelihood ratio property, under which decision procedures have a simple threshold form. In this setting, I show that decisions of Bayesian regulators converge to hypothesis testing rules as they become more pessimistic about the pool of potential proposals. Section 5 extends the problem to allow proponents of innovations to be uncertain about their value prior to collecting evidence. I show that if the signal is normally distributed, hypothesis testing rules remain admissible and minimax. Section 6 deals with strategic choice of the cost and precision of evidence.

2 Difficulties of rationalizing the hypothesis testing criterion

Though it is widely used and intuitively appealing to many researchers, the classical hypothesis testing criterion is difficult to motivate in decision problems. When testing is used to choose between two alternative treatments, which is the focus of this paper, hypothesis testing criterion raises two questions. First, why is constraining the maximum probability of Type I errors lexicographically more important than minimizing the chance of Type II errors? While one type of errors could be more important than the other, typical decision criteria call for weighted consideration of both rather than just a constraint on one. Second, why only the probability of Type I errors is considered and not their magnitude? It would seem that the probability of mistakenly choosing the wrong treatment should be more important if there is a large difference in the effectiveness of the two, and less important if they are almost equally effective. This paper proposes an answer to both of these questions.

Statistical decision problems have been analyzed as *games against nature* starting with Wald (1950) and Savage (1954). First, nature "picks" the unknown parameter value $\theta \in \Theta$. In this paper, θ will refer to the net effect of a new treatment, with the net effect of the status quo treatment normalized to zero. Then, statistical data $X \in \mathcal{X}$ is randomly drawn from probability distribution $F(X;\theta)$ that depends on the parameter θ. The statistician then makes a decision $\delta(X)$, possibly randomized, based on observed data. In case of hypothesis testing, the decision is binary: $\delta(X) = 1$ if the alternative treatment is accepted, $\delta(X) = 0$ if the status quo is chosen. The performance of the decision rule under each possible parameter value could be summarized by the expected loss

$$E_{F(X;\theta)} L(\delta(X), \theta). \tag{1}$$

Usually, no decision rule minimizes (1) for all θ. The choice of δ then depends on the criterion used to deal with ambiguity in how nature picks θ. The statistician could place a subjective prior distribution μ on Θ and minimize subjective expected loss $E_{\mu(\theta)} E_{F(X;\theta)} L(\delta(X), \theta)$. Alternatively, the statistician could look a decision rule that performs uniformly well for all θ, for example, minimizing $\sup_{\theta \in \Theta} E_{F(X;\theta)} L(\delta(X), \theta)$ (minimax criterion). Stoye (2011) provides an extensive overview of various uniform criteria and their axiomatic properties.

Discussions of hypothesis testing sometimes invoke a 1–K loss function that penalizes all

Type I errors by K points and all Type II errors by 1 point:

$$\begin{aligned} L(\delta,\theta) = {} & \delta\cdot K\cdot \mathrm{I}(\theta\in\Theta_0) \\ & + (1-\delta)\cdot \mathrm{I}(\theta\notin\Theta_0). \end{aligned} \tag{2}$$

In some problems, hypothesis tests with size $\alpha = \frac{1}{K+1}$ are minimax with this loss function. Tests with 5% significance level, for example, would then be minimax for a loss function that places 19 times more weight on any Type I error than on any Type II error. Loss function (2) does not provide a good rationalization for the use of classical hypothesis testing criterion in treatment choice. While minimax decision rules with this loss function (or Bayesian decision rules with various priors) could coincide with hypothesis testing rules, they do not motivate the lexicographic hypothesis testing criterion itself. The Type I/II error loss ratios corresponding to typical test levels (K=19 for 5% and K=99 for 1%) may seem rather large for many problems. Furthermore, this loss function ignores the substantive magnitude of committed errors, assigning the same penalty for mistakenly approving either a treatment that is only infinitesimally worse or one that is gravely inferior.

A number of recent papers considered treatment choice as a statistical decision problem using regret loss function, which penalizes both Type I and Type II errors proportionately to the magnitude of the difference in treatment effects $|\theta|$:

$$\begin{aligned} L(\delta,\theta) = {} & \delta\cdot |\theta|\cdot \mathrm{I}(\theta<0) \\ & + (1-\delta)\cdot |\theta|\cdot \mathrm{I}(\theta>0). \end{aligned} \tag{3}$$

The optimal decision rules under both minimax and average risk (flat prior) decision criteria are essentially *empirical success rules* (Manski, 2004, 2005) that prescribe choosing the treatment that appears to be more successful in trials, whether by a small or by a wide margin (see also Stoye, 2009; Hirano and Porter, 2009; Schlag, 2007). These decision rules are comparable to one-sided hypothesis tests with 50% level, rather than the conventional test levels of 5% or 1%.

The regret loss function could be easily modified to place greater weight on losses from Type I errors, similarly to (2):

$$\begin{aligned} L(\delta,\theta) = {} & \delta\cdot K\cdot |\theta|\cdot \mathrm{I}(\theta<0) \\ & + (1-\delta)\cdot |\theta|\cdot \mathrm{I}(\theta>0). \end{aligned} \tag{4}$$

Minimax and average risk decision rules with this asymmetric loss function are similar to hypothesis testing rules in that innovations are approved only if the estimate of θ exceed some multiple of its standard error (Tetenov, 2012; Hirano and Porter, 2009). However, to obtain decision rules comparable to tests with conventional levels, the asymmetry factor K has to be much greater than with loss function (2). The difference is due to accounting for the magnitude of errors in the loss function, which makes mistakes for $\theta \approx 0$ almost irrelevant to the analysis. One-sided tests with 5% level are minimax optimal if K=102, while 1% tests are optimal if K=970 (Tetenov, 2012). In contrast, loss aversion coefficient of K=3 would lead to a 34% one-sided test. While loss aversion seems like an intuitive way to motivate the asymmetry of conventional hypothesis tests, it cannot easily rationalize tests with levels similar to conventional levels and it cannot rationalize the hypothesis testing criterion itself.

3 Screening proponents through hypothesis testing

This section illustrates how the hypothesis testing could be optimal for regulatory approval under the minimax criterion. The results in this section do not rely on any statistical properties of the data generating process. Next section will impose the monotone likelihood ratio property on the distribution of clinical trial data, which allows a richer theoretical analysis (including prior beliefs about proponents for the regulator and allowing the proponents to be uncertain about the quality of their innovations).

There are two parties: the *proponent* of an innovation and the *regulator*. The regulator has to decide whether to accept or reject the proposed innovation. The quality of the innovation, which determines the parties' payoffs if it is accepted by the regulator, is $\theta \in \Theta$. The proponent knows θ, while the regulator does not (I analyze what happens if the proponent is uncertain about the value of θ in Section 5).

If the regulator accepts the innovation, the proponent gets a payoff of $b(\theta)$ and the regulator gets a payoff of $v(\theta)$. The payoff to both parties is zero if the regulator rejects the innovation. The function $b(\theta)$ is known to the regulator. In the simple case when the proponent knows θ with certainty, we could assume $b(\theta) > 0$ for all $\theta \in \Theta$ ($b(\theta) < 0$ will be allowed in Section 5 dealing with uncertain proponents). Thus any proponent wants the regulator to approve his innovation.

Let $\Theta_0 = \{\theta : v(\theta) < 0\}$ be the set of innovations that are detrimental for the regulator to accept (the null hypothesis) and $\Theta_1 = \{\theta : v(\theta) \geq 0\}$ the set of good innovations (the alternative hypothesis).

To convince the regulator that $\theta \in \Theta_1$, the proponent could conduct a trial of the innovation that costs $c > 0$ and generates a sufficient statistic $X \in \mathcal{X}$ with probability distribution $F(X;\theta)$. The data generating process $F(X;\theta)$ is known to both parties. The results of the trial then could be provided to the regulator. The full trial cost c is sunk before any data X is realized and the amount c is known to the regulator.

The regulator commits to using a *statistical decision rule* $\delta : \mathcal{X} \to [0,1]$, where $\delta(X)$ denotes the probability with which the regulator will accept the innovation when the realized outcome of the trial is X. If the regulator commits to systematically use a publicly known decision rule, then she does not need to communicate with potential proponents until they decide to submit their proposals with the supporting data X and proof that cost c was sunk to obtain the data.

To summarize, the timing of the game is as follows. Regulator commits to an approval decision rule δ as a function of the statistic X (and of the publicly known characteristics $b(\theta)$ and c of the proponent). The proponent, knowing his type θ, chooses whether to invest c in collecting the data. If the proponent chooses to collect data, nature draws X according to the distribution $F(X;\theta)$. The proponent requests and the regulator grants approval if the realization of X is favorable ($\delta(X) = 1$). The parties' payoffs are $(\delta(X) \cdot b(\theta) - c, \delta(X) \cdot v(\theta))$ if the proponent chooses to collect evidence and $(0,0)$ otherwise.

The ex ante probability that an innovation with parameter θ will be accepted following the trial is

$$A_\delta(\theta) \equiv \int_{\mathcal{X}} \delta(X) dF(X;\theta). \tag{5}$$

Since $\delta(X) = 1$ denotes rejection of the null hypothesis by the regulator, $A_\delta(\theta)$ is the test's *power* in statistical terms. Assuming that proponents are risk-neutral, it is optimal for them to invest in a trial if its expected payoff is positive, that is, if

$$A_\delta(\theta) \geq \frac{c}{b(\theta)}. \tag{6}$$

Instead, proponents of type θ who face approval probability below this threshold will be deterred by the regulator's statistical decision rule from conducting trials.

The regulator's expected payoff when facing a proponent of type θ is

$$v(\theta) \cdot A_\delta(\theta) \cdot \mathrm{I}\left[A_\delta(\theta) \cdot b(\theta) \geq c\right]. \tag{7}$$

The regulator's payoff is zero for all θ such that $A_\delta(\theta) < \frac{c}{b(\theta)}$. Otherwise, the expected payoff has the same sign as $v(\theta)$.

To achieve the maximum feasible payoff of zero for $\theta : v(\theta) < 0$, it is sufficient for the regulator to pick any decision rule such that $A_\delta(\theta) < \frac{c}{b(\theta)}$. To achieve the same objective in a game against nature (in which the decision to conduct trials is not strategic), the regulator would have to pick a decision rule with $A_\delta(\theta) = 0$. Only decision rules that reject innovations for all realizations of the data are minimax in a game against nature, which has led Manski (2004) to question the usefulness of the minimax decision criterion. In contrast, strategic entry decision by the proponent allows the regulator to employ non-degenerate decision rules even under the pessimistic minimax criterion.

Proposition 1 *Any testing decision rule δ that controls size*

$$A_\delta(\theta) < \frac{c}{b(\theta)} \textit{ for all } \theta \in \Theta_0 \tag{8}$$

is minimax for the regulator with respect to proponent type θ.

Proof. Minimax payoff for the regulator cannot be greater than zero, since zero is the highest possible payoff for $\theta \in \Theta_0$. If δ satisfies (8), then the regulator's payoff equals zero for all $\theta \in \Theta_0$. The regulator's payoff is always non-negative for $\theta \in \Theta_1$, hence δ is minimax. ■

If the proponent's payoff does not depend on θ (i.e., $b(\theta) = b$), then condition (8) in Proposition 1 simplifies to

$$A_\delta(\theta) < \frac{c}{b} \text{ for all } \theta \in \Theta_0, \tag{9}$$

which is essentially a typical statistical test size control condition, with the requisite size of the test determined by economic parameters b and c.

There could be many alternative minimax testing decision rules, some of which could be *inadmissible* (weakly or strongly dominated by other decision rules). More structure has to be placed on the data distribution $F(X;\theta)$ to determine what decision rules are admissible. In the rest of the paper, I will imposes the monotone likelihood ratio property on F, which leads to a very strong characterization of admissible decision rules.

Minimax criterion is a conservative decision criterion for choice under ambiguity. In this case, the regulator's ambiguity about the distribution of θ among potential proponents. In the next section I will consider how minimax decision rules compare to optimal decision rules for regulators who could place a prior distribution on proponent type θ and show that minimax decision rules could be seen as a limit case as regulator places prior probability on $\theta \in \Theta_0$ approaching 1.

Minimax decision rules will be optimal (or nearly optimal) if the regulator believes that the supply of bad innovations (proponents with $\theta \in \Theta_0$) is almost bottomless compared to the supply of good ones. In many contexts, an unlimited number of potential proposals with $v(\theta) < 0$ could be generated almost effortlessly, so the regulator may have a reason to be very pessimistic. The relevant distribution of proponent types is the distribution of *potential* proponents, which is difficult to observe in practice. If some kind of deterrent policy is already in place (which is the case in all fields using some form of hypothesis testing), the distribution of deterred proponents is completely hidden from the regulator.

4 Proponents with precise knowledge of θ

To better understand the properties of hypothesis testing, it is useful to put some additional structure on the problem. In this section I will consider the statistical approval problem with real-valued parameter $\theta \in \Theta \subseteq \mathbb{R}$, which describes the value of the innovation to the regulator: $v(\theta) = \theta$.

I will also assume that the sufficient statistic X of the collected data has distribution functions $F(X;\theta)$ with density $f(x;\theta)$ that satisfy the *monotone likelihood ratio property:*

$$x_1 > x_2, \theta_1 > \theta_2 \Rightarrow f(x_1;\theta_1) \cdot f(x_2;\theta_2) \geq f(x_1;\theta_2) \cdot f(x_2;\theta_1). \tag{10}$$

Also, let the cumulative distributions $F(X \leq t;\theta)$ be continuous in θ for all t, satisfying $f(x;\theta) > 0$ for all $x \in \mathcal{X}$ and $F(X;\theta) \neq F(X;\theta')$ for $\theta \neq \theta'$. A leading example is $X \sim \mathcal{N}(\theta, \sigma^2)$ for some known $\sigma > 0$.

Let the innovator's benefit from the acceptance decision $b(\theta)$ be continuous and non-decreasing in θ, with $b(0) > 0$, meaning that proponents are interested in getting approval for a range of ideas with $\theta < 0$. The value that the proponent would obtain from acceptance of an idea with zero effect $b\,(0)$ is the only parameter of this function that needs to be known to the regulator.

In this setting, it is sufficient for the regulator to consider *monotone decision rules* of the form:

$$\delta_{T,\lambda}(X) = \begin{cases} 0 & \text{for } X < T, \\ 1-\lambda & \text{for } X = T, \\ 1 & \text{for } X > T. \end{cases} \tag{11}$$

Lemma 3 in Karlin and Rubin (1956) establishes that for any decision rule $\delta : X \to [0,1]$, there exists a unique monotone decision rule δ' which yields higher approval probability for all good innovations and lower approval probability for bad ones:

$$\begin{aligned} &A_{\delta'}(\theta) \le A_{\delta}(\theta) \text{ for all } \theta \le 0 \text{ and} \\ &A_{\delta'}(\theta) \ge A_{\delta}(\theta) \text{ for all } \theta \ge 0. \end{aligned} \tag{12}$$

The proponent's decision to conduct a trial is then higher under δ' for positive θ and lower for negative θ:

$$\begin{aligned} &\mathrm{I}\left[A_{\delta'}(\theta)\cdot b(\theta) \ge c\right] \le \mathrm{I}\left[A_{\delta}(\theta)\cdot b(\theta) \ge c\right] \ \text{ for all } \theta \le 0 \text{ and} \\ &\mathrm{I}\left[A_{\delta'}(\theta)\cdot b(\theta) \ge c\right] \ge \mathrm{I}\left[A_{\delta}(\theta)\cdot b(\theta) \ge c\right] \ \text{ for all } \theta \ge 0. \end{aligned} \tag{13}$$

Therefore, the regulator's payoff (7) from δ' is at least as large as the payoff from δ for all θ.

The acceptance probability $A_\delta(\theta)$ of monotone decision rules is decreasing lexicographically in (T,λ) for every θ. There is a monotone decision rule δ^* for which the acceptance probability at $\theta = 0$ is exactly:

$$A_{\delta^*}(0) = \frac{c}{b(0)}. \tag{14}$$

This decision rule is the same as a one-sided hypothesis test of $H_0 : \theta \le 0$ with size $\frac{c}{b(\theta)}$. If $X \sim \mathcal{N}(\theta, \sigma^2)$ with known σ, then

$$\delta^*(X) = \mathrm{I}\left[X > \sigma\Phi^{-1}\left(\frac{c}{b\,(0)}\right)\right]. \tag{15}$$

Due to the monotone likelihood ratio property, the acceptance probability of monotone decision rules is increasing in θ, so innovations with $\theta < 0$ all face acceptance probability below $\frac{c}{b(0)}$ and investing in trials is not optimal for their proponents. All innovations with $\theta > 0$ face higher acceptance probability and their proponents will find it optimal to invest in trials. The following proposition shows that δ^* is an admissible and minimax decision rule. All monotone decision rules with higher acceptance probability at $\theta = 0$ are not minimax because they make it profitable for proponents with some $\theta < 0$ to conduct trials and the regulator faces expected payoff below zero (which is the minimax value) for that type of proponents. Decision rules with lower acceptance probability at $\theta = 0$ are minimax but are not admissible because δ^* yields the same zero payoff for the regulator for all $\theta \leq 0$, but provides higher acceptance probability, hence higher payoffs, for $\theta > 0$.

Proposition 2 *Monotone decision rules δ^* satisfying (14) are minimax and admissible.*

Proof. Minimum expected payoff (over θ) for the regulator cannot be higher than zero, which is the payoff for $\theta = 0$ no matter which decision rule the regulator chooses. To show that δ^* is minimax, we just need to show that it yields nonnegative expected payoff to the regulator for all θ.

It is not optimal for potential proponents with $\theta < 0$ to invest in collecting evidence because

$$A_{\delta^*}(\theta)b(\theta) < A_{\delta^*}(0)b(0) = c. \tag{16}$$

The inequality follows from monotonicity of $A_{\delta^*}(\cdot)$ and $b(\cdot)$. The regulator's payoff from facing this type of proponent is zero and cannot be improved by any decision rule.

For proponents with $\theta \geq 0$, the regulator's payoff is nonnegative for any decision rule. Hence δ^* is minimax.

Decision rule δ^* is admissible if there does not exist a decision rule δ' which yields strictly better payoff for some value of θ and is at least as good as δ^* for all values of θ. If any decision rule dominates δ^*, then there is also a monotone decision rule δ' that dominates δ^*.

If $A_{\delta'}(0) > \frac{c}{b(0)}$, then, by continuity, $A_{\delta'}(\theta')b(\theta') > c$ also for some $\theta' < 0$. This makes it profitable for the proponents of type θ' to invest in trials. Then, the regulator's expected payoff is below zero at θ', hence δ' cannot dominate δ^*.

If $A_{\delta'}(0) < \frac{c}{b(0)}$, then also $A_{\delta'}(\theta')b(\theta') < c$ for some $\theta' > 0$, hence these proponents will not find it optimal to invest in trials and the regulator's expected payoff at θ' will be zero. Under δ^*, all proponents with $\theta > 0$ will find it profitable to conduct trials, yielding strictly positive expected payoff to the regulator. Hence δ' cannot dominate δ^*. ■

4.1 Decision rules for Bayesian regulators

The idea of potentially unlimited entry of proponents with bad ideas may seem overstretched. Below I show how hypothesis testing rules with test level (14) relate to decision rules of a regulator who has a prior distribution over the types of potential proponents. First, I show that the hypothesis tests with level (14) always set a higher threshold than Bayesian rules for approval of innovations. Second, I show that the hypothesis test rule is a limit of Bayesian procedures adopted by regulators who assume higher and higher proportion of potential proponents to have bad ideas.

I will denote the proponent's type simply by θ and the regulator's prior over proponents' types by $Q(\theta)$ with density $q(\theta)$. For simplicity, let Q be atomless.

The regulator's expected payoff from using decision rule δ equals

$$V(\delta) \equiv \int_{\Theta} \theta A_{\delta}(\theta) \cdot \mathrm{I}\left[A_{\delta}(\theta) b(\theta) > c\right] dQ(\theta). \tag{17}$$

It is sufficient to consider maximizing (17) over the set of monotone decision rules (11). To simplify exposition, assume that $F(X;\theta)$ is also atomless, so that randomization λ on the threshold in monotone decision rules could be ignored. The regulator's maximization problem then reduces to a one-dimensional problem of finding an optimal threshold T:

$$\max_{T} \int_{\bar{\theta}(T)}^{+\infty} \theta A_{\delta_T}(\theta) dQ(\theta). \tag{18}$$

Since $A_{\delta_T}(\theta) b(\theta)$ is increasing in θ for all θ s.t. $b(\theta) > 0$, the range of θ for which $A_{\delta_T}(\theta) b(\theta) > c$ is an interval $(\bar{\theta}(T), +\infty)$, where $\bar{\theta}(T) \equiv \inf\{\theta : A_{\delta_T}(\theta) b(\theta) > c\}$ is the lowest value of θ at which it is profitable for proponents to collect data.

The first result is that for Bayesian regulators it is generally optimal to set the threshold lower than for the hypothesis test (14). While a fraction of proponents with bad ideas ($\theta < 0$) will find it optimal to try (and hence gain approval with some probability), this loss is offset by higher probability of success for proponents of good ideas.

Proposition 3 *If the regulator believes that a positive measure of potential proponents have good ideas ($Q(\theta > 0) > 0$), then the optimal decision rule δ_T will have a lower threshold than the hypothesis test rule (14).*

Proof. The hypothesis test threshold T^* is constructed so that $A_{\delta_{T^*}}(0)b(0) = c$, so $\bar{\theta}(T^*) = 0$ (only proponents with $\theta > 0$ find it profitable to collect data). Consider the derivative of (18) with respect to T:

$$\frac{d}{dT}\int_{\bar{\theta}(T)}^{+\infty}\theta A_{\delta_T}(\theta)\,dQ(\theta)=-\frac{d\bar{\theta}(T)}{dT}\cdot\bar{\theta}(T)\,A_{\delta_T}\left(\bar{\theta}(T)\right)+\int_{\bar{\theta}(T)}^{+\infty}\theta\frac{dA_{\delta_T}(\theta)}{dT}q(\theta)\,d\theta. \tag{19}$$

The first effect is that increasing T makes it unprofitable for proponents with θ slightly above $\bar{\theta}(T)$ to collect data, this effect is positive for $T<T^*$ and negative for $T>T^*$ (because then proponents of good innovations are deterred). The second effect is that increasing T reduces probability of approval for all proponents with $\theta>\bar{\theta}(T)$. Evaluated at $T=T^*$, the first term equals zero because $\bar{\theta}(T^*)=0$ and the second term is negative because $dA_{\delta_T}(\theta)/dT<0$, $\theta>0$ and $q(\theta)>0$ on a set of positive measure, so

$$\left.\frac{d}{dT}V(\delta_T)\right|_{T=T^*}=\int_0^{+\infty}\theta\frac{dA_{\delta_T}(\theta)}{dT}q(\theta)\,d\theta<0. \tag{20}$$

Hence optimal threshold T is always smaller than T^*. ■

The next result shows that if the regulator puts more and more prior mass on $\theta<0$, the thresholds of optimal decision rules converge to the hypothesis test threshold T^*.

Proposition 4 *Let Q be a probability measure on real line with density q and finite mean, such that $Q([0,+\infty))>0$ and $q(\theta)>0$ on $[\varepsilon,0]$ for some $\varepsilon<0$. Denote $Q^+(A)=Q(A\cap[0,+\infty))$ and $Q^+(A)=Q(A\cap(-\infty,0))$. Let $\{a_n\}$ be an increasing sequence of numbers $a_n\in(0,1)$, $a_n\to 1$. Define a sequence of probability measures $Q_n=\frac{a_n}{Q((-\infty,0))}Q^-+\frac{1-a_n}{Q([0,+\infty))}Q^+$ and the sequence of optimal decision rule thresholds T_n that maximize (18) with respect to Q_n. Then $T_n\to T^*$.*

Proof. It follows from Proposition 3 that $T_n<T^*$ for all n. I proceed to show that given any $\bar{T}<T^*$, $T_n>\bar{T}$ for sufficiently large n.

For every n, $V(\delta_{T_n})>V(\delta_{T^*})>0$. On the other hand, I will show that for sufficiently large n, $V(\delta_T)<0$ for all $T\le\bar{T}$, hence $T_n>\bar{T}$. Since $T\le\bar{T}<T^*$, $\bar{\theta}(T)\le\bar{\theta}\left(\bar{T}\right)<0$ and $A_{\delta_T}(\theta)\ge A_{\delta_{\bar{T}}}(\theta)$,

$$\begin{aligned}V(\delta_T)&=\int_{\bar{\theta}(T)}^{0}\theta A_{\delta_T}(\theta)\,q_n(\theta)\,d\theta+\int_0^{+\infty}\theta A_{\delta_T}(\theta)\,q_n(\theta)\,d\theta\\&=\frac{a_n}{Q((-\infty,0))}\int_{\bar{\theta}(T)}^{0}\theta A_{\delta_T}(\theta)\,q(\theta)\,d\theta+\frac{1-a_n}{Q([0,+\infty))}\int_0^{+\infty}\theta A_{\delta_T}(\theta)\,q(\theta)\,d\theta\\&\le\frac{a_n}{Q((-\infty,0))}\int_{\bar{\theta}(\bar{T})}^{0}\theta A_{\delta_{\bar{T}}}(\theta)\,q(\theta)\,d\theta+\frac{1-a_n}{Q([0,+\infty))}\int_0^{+\infty}\theta q_n(\theta)\,d\theta.\end{aligned} \tag{21}$$

The first integral is negative (because it is taken over $\theta < 0$) and the second is positive. For sufficiently large n, their weighted sum is negative, which implies that $V(\delta_T) < 0$ for all $T \leq \bar{T}$.

■

This proposition shows that the hypothesis testing rule with level $\frac{c}{b(0)}$ could be interpreted as an approximation of Bayesian decision rules that would be taken by regulators sufficiently pessimistic about the pool of potential proponents seeking approval for their "innovations."

5 Proponents ex ante uncertain about the value of their proposals

Proponents may have some information about the value of their innovation, but not enough to know θ exactly. The proponent's type in this case is π - the probability distribution that the proponent places on θ prior to collecting evidence. Here $\pi \in \Delta$ could be any probability distribution on $\Theta \subset \mathbb{R}$ with finite mean. The subset of degenerate distributions (beliefs of proponents certain about θ) will be denoted by Δ_0.

It is profitable for the proponent to invest in collecting evidence if

$$\int_{\Theta} b(\theta) A_\delta(\theta) d\pi(\theta) - c > 0 \tag{22}$$

This payoff depends on the shape of the prior π in a nontrivial fashion, since $A_\delta(\theta)$ is nonlinear.

The regulator also needs to take into account that proponents may not want to seek approval for their innovations after observing the data X. This could happen if the proponent's posterior beliefs $\pi(\theta|X)$ place enough weight on θ for which $b(\theta)$ is negative. To accommodate this possibility in notation, denote the proponent's optimal decision whether to seek approval after observing the data by the function $\eta_\pi(X) = \mathrm{I}\left[\int_\Theta b(\theta) d\pi(\theta|X)\right]$. The ex ante probability of approval then becomes

$$A_\delta(\theta) = \int_{\mathcal{X}} \delta(X) \eta_\pi(X) dF(X;\theta) \tag{23}$$

It turns out that threshold hypothesis testing rule (14) with test level $\frac{c}{b(0)}$ is also minimax and admissible in this setting if the proponent's benefit from approval $b(\theta)$ is concave in θ and $\frac{A'_\delta(\theta)}{A_\delta(\theta)}$ is non-increasing in θ.

Examples of distributions for which $\frac{A'_\delta(\theta)}{A_\delta(\theta)}$ is non-increasing in θ include the family of normal distributions $X \sim \mathcal{N}(\theta, \sigma^2)$ with known variance and the family of exponential distributions with means $\mu_0 + \theta$.

The following proposition shows that collecting evidence is only optimal for proponents with beliefs for which the regulator would also expect to receive positive payoffs.

Proposition 5 *Suppose that the function $b(\theta)$ is non-decreasing and concave in θ, hypothesis test rule δ^* given in (14) is used by the regulator, and $\frac{A'_{\delta^*}(\theta)}{A_{\delta^*}(\theta)}$ is non-increasing in θ. Whenever the regulator's expected payoff would be negative if proponents of type π collected data*

$$\int_\Theta \theta A_{\delta^*}(\theta)\, d\pi(\theta) < 0, \tag{24}$$

it is not optimal for proponent to collect data:

$$\int_\Theta b(\theta)\, A_{\delta^*}(\theta)\, d\pi(\theta) - c < 0. \tag{25}$$

Proof. See Appendix. ■

It follows from Proposition 5, that the regulator faces non-negative expected payoff from all proponent types π, hence δ^* is minimax.

Proposition 6 *Suppose that the function $b(\theta)$ is non-decreasing, continuous, and concave in θ, and that $\frac{A'_\delta(\theta)}{A_\delta(\theta)}$ is non-increasing in θ for threshold decision rules δ. Then the hypothesis testing rule δ^* in (15) is minimax and admissible with respect to $\pi \in \Delta$.*

Proof. For δ^* to be inadmissible, there must be a different monotone decision rule δ that yields strictly better payoffs to the regulator for some π and does not yield worse payoffs for any π. Any other monotone decision rule $\delta(X)$, however, either has $A_\delta\left(\bar{\theta}\right) < A_{\delta^*}\left(\bar{\theta}\right)$ for all $\bar{\theta} > 0$, in which case it yields lower expected payoff to the regulator for π that puts probability one on $\bar{\theta}$, or $A_\delta(0) > \frac{c}{b(0)}$, in which case it yields negative payoffs to the regulator for some certain type of proponent with $\theta < 0$.

It follows from Proposition 5 that the regulator's payoffs are non-negative for any $\pi \in \Delta$ and they equal zero for π_0 that puts mass one on $\theta = 0$. No decision rule could yield a higher minimum, since all decision rules yield payoff of zero for π_0, hence δ^* is minimax. ■

One-sided hypothesis testing rules with level $\frac{c}{b(0)}$ have attractive decision-theoretic properties if the regulator cannot place more precise restrictions on the distribution of different proponent types, at least if the evidence is normally distributed and the proponent's benefit from approval is concave in θ. It is unclear under what kinds of more general conditions the same result might hold.

The results of Propositions 5 and 6 do not apply if the benefit function $b(\theta)$ is not concave. The following example illustrates that proponents with non-concave payoff functions $b(\theta)$ may find it profitable to test innovations that yield a negative expected payoff to the regulator.

Example 7 *Suppose that the proponent's benefit from approval is $b(-1) = 0$, $b(0) = 1$, and $b(1) = 10$. Let $F(X; \theta) = \mathcal{N}(\theta, 1)$ and let $c = \frac{1}{2}$. Then the hypothesis test rule (15) should have size $\frac{1}{2}$ and is simply $\delta^*(X) = \mathrm{I}[X > 0]$. The probability of approval as a function of θ equals $A_{\delta^*}(\theta) = F(X < 0; \theta) = \Phi(\theta)$.*

If the proponent's prior beliefs place probabilities $\pi(\theta = -1) = .9$ and $\pi(\theta = 1) = .1$, then the expected payoff from testing for the proponent is positive

$$\sum_{\theta \in \{-1,1\}} b(\theta) A_{\delta*}(\theta) \pi(\theta) - c = 0 + 10 * \Phi(1) * 0.1 - 0.5 \approx .341, \tag{26}$$

while the expected payoff to the regulator is negative

$$\sum_{\theta \in \{-1,1\}} \theta A_{\delta*}(\theta) \pi(\theta) = -1 * \Phi(-1) * .9 + 1 * \Phi(1) * .1 \approx -0.059. \tag{27}$$

6 Endogenous choice of testing cost and precision

So far, the choice of testing cost c and the data distribution $F(X; \theta)$ (for example, sample size) was fixed and not included as a step in the game. Even when proponents could choose $(c, F(X, \theta))$, hypothesis test rules with the proposed level $\frac{c}{b(0)}$ remain effective as a deterrent for proponents with $\theta < 0$, as long as the test level is based on their chosen trial costs. However, it may be optimal for the regulator to make the test stricter for some (c, F) to persuade proponents to choose other values of (c, F) that would yield higher expected payoff for the regulator.

I first illustrate that if all proponents were ex ante certain about the value of θ and could freely choose how much to spend on collecting evidence, the regulator would push all the proponents towards choosing $c = b(0)$. This effectively replaces statistical signaling with pure monetary signaling of the proponent's type.

Example 8 *Suppose that $\pi \in \Delta_0$ and that $b(\theta) > b(0)$ for all $\theta > 0$ and $b(\theta) < b(0)$ for all $\theta < 0$. Let $\delta^*_{(c,F)}$ be a decision rule that satisfies condition (8) for each choice of (c, F). If the* *proponent's choice set of* (c, F) *includes at least one choice with* $c = b(0)$*, then an alternative decision rule*

$$\begin{aligned} &\delta(X) = 1 \text{ if } c = b(0), \\ &\delta(X) = 0 \text{ if } c \neq b(0). \end{aligned} \tag{28}$$

weakly dominates $\delta^*_{(c,F)}$.

For all innovators with $\theta > 0$ *it remains profitable to conduct testing and the approval rate for their innovations is 1, which yields the highest possible expected payoff to the regulator. On the other hand, attempting "data collection" remains unprofitable for innovators with* $\theta < 0$.

If $b(\theta) = b(0)$, then almost the same effect could be obtained by decision rules that only slightly depend on the statistical signal:

$$\begin{aligned} &\delta(X) = (1-\varepsilon) + \varepsilon\delta^*_{(c,F)}(X) && \text{if } c = b(0), \\ &\delta(X) = 0 && \text{if } c \neq b(0) \end{aligned} \tag{29}$$

for $\varepsilon \to 0$. The approval rate is smaller than one for $\theta < 0$, which still makes it unprofitable for innovators with $\theta < 0$ to conduct experiments. But for all $\theta > 0$, the approval rate is closer to 1 then the approval rate under $\delta^*_{(c,F)}$ for any $\varepsilon \in (0, 1)$.

Even when proponents could choose (c, σ), the results of Proposition 5 apply, hence $\delta^*(X)$ is a minimax decision rule if $b(\cdot)$ is linear and $\pi \in \Delta$. The problem is that we do not know whether it is admissible. For example, the decision rule $\delta_0(X) = 0$ that rejects all innovations is also minimax, since it always yields a payoff of zero to the regulator. It is not admissible because the hypothesis testing rule yields strictly higher payoffs to the regulator for some types of proponents and never less than zero. Similarly, we cannot yet rule out that when proponents could choose (c, σ) there may be decision rules that dominate the hypothesis testing rule δ^*.

The following proposition is a small step towards resolving this problem. I show that if the proponent could choose between two pairs (c_1, σ_1) and (c_2, σ_2) with higher costs corresponding to more precise evidence, and if all types of proponents $\pi \in \Delta$ are allowed, then the decision rule δ^* is admissible.

Proposition 9 *Suppose that* $X \sim \mathcal{N}(\theta, \sigma^2)$, $\pi \in \Delta$ *and* $b(\theta) = b$ *for all* θ*. If the proponent could choose between two pairs* (c_1, σ_1) *and* (c_2, σ_2) *with* $\frac{1}{2}b > c_1 > c_2$ *and* $\sigma_1 < \sigma_2$*, then the decision rule*

$$\delta^*(X; c, \sigma) = \mathrm{I}\left[X > \sigma\Phi^{-1}\left(\frac{c}{b}\right)\right] \text{ for each } c \tag{30}$$

is admissible.

Proof. See Appendix. ■

7 Conclusion

The paper presented a novel rationalization for the use of hypothesis testing rules in statistical treatment choice. The probability of Type I errors for innovations with negative effects has to be contained to deter potential proponents from flooding the regulator with random policy proposals. The test level is determined by the ratio of the proponent's testing costs over her expected benefits from the proposal's approval. In this setting, hypothesis testing rules turn out to be admissible and minimax decision rules for the regulator with respect to the proponent's ex ante beliefs about the quality of the innovation. They can also be seen as a limit of testing rules adopted by Bayesian regulators who are sufficiently pessimistic about the proportion of opportunistic proposals.

8 Appendix

Proof of Proposition 5. The proponent's approval decision rule always has the form $\eta(X) = \mathrm{I}\left[X > T_P\right], T_P \geq -\infty$ due to the monotone likelihood ratio property of $F(X;\theta)$. Then the joint decision rule (probability that both the regulator and the proponent want to approve the innovation after observing X) is also a threshold decision rule

$$\tilde{\delta}(X) = \delta^{*}(X)\eta(X) = \mathrm{I}\left[X > T\right] \tag{31}$$

and could be expressed as a one-sided hypothesis testing rule with level $A_{\tilde{\delta}}(0) = \frac{\tilde{c}}{b(0)}$ for some $\tilde{c} \leq c$.

Function $b(\theta)$ can be bounded from above by a linear function that passes through $b(0)$ because it is concave:

$$\tilde{b}(\theta) = b(0) + \beta\theta \geq b(\theta). \tag{32}$$

Since $b(\theta)$ is non-decreasing, $\beta \geq 0$.

The innovator's expected payoff from collecting evidence then equals

$$(\tilde{c}-c)+\int[b(\theta)A_{\tilde{\delta}}(\theta)-\tilde{c}]d\pi(\theta)\leq(\tilde{c}-c)+\int[\tilde{b}(\theta)A_{\tilde{\delta}}(\theta)-\tilde{c}]d\pi(\theta). \tag{33}$$

The integrand $[\tilde{b}(\theta)A_{\tilde{\delta}}(\theta)-\tilde{c}]$ is zero at $\theta=0$, positive for $\theta>0$ and negative for $\theta<0$. The same is true for $\theta A_{\tilde{\delta}}(\theta)$. Hence the ratio

$$r(\theta)=\frac{\tilde{b}(\theta)A_{\tilde{\delta}}(\theta)-\tilde{c}}{\theta A_{\tilde{\delta}}(\theta)} \tag{34}$$

is positive for all $\theta\neq0$. While it is not defined at $\theta=0$, it has a well-defined limit from both sides as $\theta\to0$, by L'Hopital's rule

$$\begin{aligned}
r(0) &= \lim_{\theta\to0}\frac{\tilde{b}(\theta)A_{\tilde{\delta}}(\theta)-\tilde{c}}{\theta A_{\tilde{\delta}}(\theta)}=\frac{\lim_{\theta\to0}[\tilde{b}(\theta)A_{\tilde{\delta}}(\theta)-\tilde{c}]'}{\lim_{\theta\to0}[\theta A_{\tilde{\delta}}(\theta)]'}= && (35)\\
&= \frac{\lim_{\theta\to0}\left[\beta A_{\tilde{\delta}}(\theta)+[b(0)+\beta\theta]A'_{\tilde{\delta}}(\theta)\right]}{\lim_{\theta\to0}\left[A_{\tilde{\delta}}(\theta)+\theta A'_{\tilde{\delta}}(\theta)\right]}\\
&= \frac{\beta A_{\tilde{\delta}}(0)+b(0)A'_{\tilde{\delta}}(0)}{A_{\tilde{\delta}}(0)} && (36)\\
&= \beta+b(0)\frac{A'_{\tilde{\delta}}(0)}{A_{\tilde{\delta}}(0)},
\end{aligned}$$

$r(0)>0$. I will show later in the proof that $r(\theta)>r(0)$ for all $\theta<0$ and that $r(\theta)<r(0)$ for all $\theta>0$.

It follows that

$$\int_0^{+\infty}\left[\tilde{b}(\theta)A_{\tilde{\delta}}(\theta)-\tilde{c}\right]d\pi(\theta)=\int_0^{+\infty}r(\theta)\,\theta A_{\tilde{\delta}}(\theta)\,d\pi(\theta)\leq r(0)\int_0^{+\infty}\theta A_{\tilde{\delta}}(\theta)\,d\pi(\theta) \tag{37}$$

and

$$\int_{-\infty}^{0}\left[\tilde{b}(\theta)A_{\tilde{\delta}}(\theta)-\tilde{c}\right]d\pi(\theta)=\int_{-\infty}^{0}r(\theta)\,\theta A_{\tilde{\delta}}(\theta)\,d\pi(\theta)\leq r(0)\int_{-\infty}^{0}\theta A_{\tilde{\delta}}(\theta)\,d\pi(\theta). \tag{38}$$

Adding the two inequalities yields

$$\int_{\mathbb{R}}\left[\tilde{b}(\theta)A_{\tilde{\delta}}(\theta)-\tilde{c}\right]d\pi(\theta)\leq r(0)\int_{\mathbb{R}}\theta A_{\tilde{\delta}}(\theta)\,d\pi(\theta), \tag{39}$$

hence if $\int_{\mathbb{R}} \theta A_{\tilde{\delta}}(\theta)\, d\pi(\theta) < 0$, then also

$$\int_{\mathbb{R}} [b(\theta)A_{\tilde{\delta}}(\theta) - c]d\pi(\theta) \leq \int_{\mathbb{R}} [\tilde{b}(\theta)A_{\tilde{\delta}}(\theta) - \tilde{c}]d\pi(\theta) < 0. \tag{40}$$

To conclude the proof of the proposition, it remains to be shown that $r(\theta) > r(0)$ for all $\theta < 0$ and that $r(\theta) < r(0)$ for all $\theta > 0$.

The ratio $\frac{A'_{\tilde{\delta}}(\theta)}{A_{\tilde{\delta}}(\theta)}$ is non-increasing in θ by assumption.

Take any $\theta < 0$, then for all $\theta < \bar{\theta} < 0$,

$$\frac{A'_{\tilde{\delta}}(\bar{\theta})}{A_{\tilde{\delta}}(\theta)} > \frac{A'_{\tilde{\delta}}(\bar{\theta})}{A_{\tilde{\delta}}(\bar{\theta})} > \frac{A'_{\tilde{\delta}}(0)}{A_{\tilde{\delta}}(0)}, \tag{41}$$

since $A_{\tilde{\delta}}(\theta) < A_{\tilde{\delta}}(\bar{\theta})$. Then the difference $A_{\tilde{\delta}}(0) - A_{\tilde{\delta}}(\theta)$ could be bounded from below by

$$\begin{aligned} A_{\tilde{\delta}}(0) - A_{\tilde{\delta}}(\theta) &= \int_{\theta}^{0} A'_{\tilde{\delta}}(\bar{\theta})\, d\bar{\theta} = A_{\tilde{\delta}}(\theta) \int_{\theta}^{0} \frac{A'_{\tilde{\delta}}(\bar{\theta})}{A_{\tilde{\delta}}(\theta)} d\bar{\theta} \\ &> A_{\tilde{\delta}}(\theta) \int_{\theta}^{0} \frac{A'_{\tilde{\delta}}(0)}{A_{\tilde{\delta}}(0)} d\bar{\theta} = -\theta A_{\tilde{\delta}}(\theta) \frac{A'_{\tilde{\delta}}(0)}{A_{\tilde{\delta}}(0)}. \end{aligned} \tag{42}$$

Substituting $\tilde{c} = b(0)\, A_{\tilde{\delta}}(0)$ and $b(\theta) = b(0) + \beta\theta$

$$\begin{aligned} r(\theta) - r(0) &= \frac{\tilde{b}(\theta)A_{\tilde{\delta}}(\theta) - \tilde{c}}{\theta A_{\tilde{\delta}}(\theta)} - \beta - b(0)\frac{A'_{\tilde{\delta}}(0)}{A_{\tilde{\delta}}(0)} \\ &= \frac{(b(0) + \beta\theta)\, A_{\tilde{\delta}}(\theta) - b(0)\, A_{\tilde{\delta}}(0)}{\theta A_{\tilde{\delta}}(\theta)} - \beta - b(0)\frac{A'_{\tilde{\delta}}(0)}{A_{\tilde{\delta}}(0)} \\ &= b(0)\left[\frac{A_{\tilde{\delta}}(\theta) - A_{\tilde{\delta}}(0)}{\theta A_{\tilde{\delta}}(\theta)} - \frac{A'_{\tilde{\delta}}(0)}{A_{\tilde{\delta}}(0)}\right] > 0. \end{aligned} \tag{43}$$

The proof for $\theta > 0$ is analogous. ■

Proof of Proposition 9. For δ^* to be inadmissible, there has to exist another decision rule δ that satisfies condition (8), gives the regulator payoffs at least as high as δ^* for any type π of proponent, and gives strictly higher payoffs for at least one type π. In particular, δ cannot give lower payoffs to the regulator for any π.

First, I will show that such a decision rule δ must be equal to δ^* for (c_1, σ_1). This is established by considering proponents with beliefs that place mass one on $\bar{\theta} > 0$ for $\bar{\theta}$ sufficiently close to zero. At $\theta = 0$, $A'_{\delta^*(\cdot;c_1,\sigma_1)}(\theta) > A'_{\delta^*(\cdot;c_2,\sigma_2)}(\theta)$. This implies that for a range of values of $\bar{\theta} > 0$, proponents who believe that $\theta = \bar{\theta}$ prefer to choose (c_1, σ_1) under δ^*. By the Neyman-Pearson lemma, any decision rule based on X with variance σ_1^2 different from $\delta^*(\cdot; c_1, \sigma_1)$ either violates (8) or offers lower probability of acceptance at $\bar{\theta}$, hence yields lower payoffs to the regulator. Because of the higher variance of X under the choice of (c_2, σ_2), the probability of acceptance at $\bar{\theta}$ is also lower under any decision rule $\delta(\cdot; c_2, \sigma_2)$ if the proponent is enticed to choose (c_2, σ_2) instead.

With $\delta(\cdot; c_1, \sigma_1) = \delta^*(\cdot; c_1, \sigma_1)$ fixed, I will show that also any deviations from $\delta^*(\cdot; c_2, \sigma_2)$ will either violate (8) or reduce the regulator's payoffs for some proponent type π. I will consider proponents with two point priors that place weight q on $\theta_1 < 0$ and weight $1 - q$ on $\theta_2 > 0$. In particular, choose θ_2 sufficiently high so that $A_{\delta^*(\cdot;c_2,\sigma_2)}(\theta_2) - c_2 > A_{\delta^*(\cdot;c_1,\sigma_1)}(\theta_2) - c_1$. This is possible because as $\theta \to \infty$, $A_{\delta^*(\cdot;c_2,\sigma_2)}(\theta_2) - c_2 \to 1 - c_2$, while $A_{\delta^*(\cdot;c_1,\sigma_1)}(\theta_2) - c_1 \to 1 - c_1 < 1 - c_2$.

For a given q and a decision rule $\delta(\cdot; c_2, \sigma_2)$, the proponent's expected payoff from collecting evidence with cost/precision (c_2, σ_2) equals

$$b \cdot \left[q\left(A_{\delta(\cdot;c_2,\sigma_2)}(\theta_1) - c_2\right) + (1-q)\left(A_{\delta(\cdot;c_2,\sigma_2)}(\theta_2) - c_2\right)\right]. \tag{44}$$

For δ^*, $A_{\delta^*(\cdot;c_2,\sigma_2)}(\theta_1) - c_2 < 0$, while $\left(A_{\delta^*(\cdot;c_2,\sigma_2)}(\theta_2) - c_2\right) > 0$, so it is optimal to collect evidence if $q > \bar{q}$ for some $\bar{q}$. Furthermore, $A_{\delta^*(\cdot;c_1,\sigma_1)}(\theta_1) < A_{\delta^*(\cdot;c_2,\sigma_2)}(\theta_1)$ and $A_{\delta^*(\cdot;c_1,\sigma_1)}(\theta_2) < A_{\delta^*(\cdot;c_2,\sigma_2)}(\theta_2)$, so it is not optimal for the proponent with any q to choose (c_1, σ_1) under δ^*. Under any alternative $\delta(\cdot; c_2, \sigma_2)$, $A_{\delta(\cdot;c_2,\sigma_2)}(\theta_2)$ must be lower than $A_{\delta^*(\cdot;c_2,\sigma_2)}(\theta_2)$ by the Neyman-Pearson lemma. If also $A_{\delta(\cdot;c_2,\sigma_2)}(\theta_1) \leq A_{\delta^*(\cdot;c_2,\sigma_2)}(\theta_1)$, then there is a range of values of q for which it is no longer profitable for the proponent to test the innovation under δ, while it was profitable under δ^*, hence the regulator receives a lower payoff (the regulator's payoff under δ^* was positive by Proposition 5). If instead $A_{\delta(\cdot;c_2,\sigma_2)}(\theta_1) > A_{\delta^*(\cdot;c_2,\sigma_2)}(\theta_1)$, then the regulator's payoff is lower because the probability of accepting the innovation if it is good ($\theta = \theta_2 > 0$) has been reduced, while the probability of accepting the innovation if $\theta = \theta_1 < 0$ has been increased. ■

References

[1] Cohen, J. (1994), "The earth is round ($p < .05$)," *American Psychologist,* 49, 997-1003.

[2] Hirano, K., and Porter, J.R. (2009), "Asymptotics for Statistical Treatment Rules," *Econometrica,* 77, 1683–1701.

[3] International Committee on Harmonisation (1998), *Guideline E9: Statistical Principles for Clinical Trials.*

[4] Johnson, D.H. (1999), "The Insignificance of Statistical Significance Testing," *The Journal of Wildlife Management,* 63, 763–772.

[5] Karlin, S., and Rubin, H. (1956), "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *The Annals of Mathematical Statistics,* 27, 272-299.

[6] Manski, C.F. (2004), "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica,* 72, 1221–1246.

[7] Manski, C.F. (2005), *Social Choice with Partial Knowledge of Treatment Response,* Princeton, NJ: Princeton Univ. Press.

[8] Manski, C.F. (2013), "Randomizing Regulatory Approval for Diversification and Deterrence," working paper, Northwestern University.

[9] Manski, C.F., and Tetenov, A. (2007), "Admissible Treatment Rules for a Risk-Averse Planner with Experimental Data on an Innovation," *Journal of Statistical Planning and Inference,* 137, 1998–2010.

[10] Ottaviani, M., and Henry, E. (2013), "Research and the Approval Process," working paper, Bocconi University.

[11] Savage, L.J. (1954), *The Foundations of Statistics*, New York: Wiley.

[12] Schlag, K.H. (2007), "Eleven – Designing Randomized Experiments Under Minimax Regret," working paper, European University Institute.

[13] Simon, H. A. (1945), "Statistical Tests as a Basis for "Yes-No" Choices," *Journal of the American Statistical Association*, 40, 80–84.

[14] Stoye, J. (2009), "Minimax Regret Treatment Choice with Finite Samples," *Journal of Econometrics,* 151, 70–81.

[15] Stoye, J. (2011), "Statistical Decisions under Ambiguity," *Theory and Decision,* 70, 129–148.

[16] Tetenov, A. (2012), "Statistical Treatment Choice Based on Asymmetric Minimax Regret Criteria," *Journal of Econometrics*, 166, 157-165.

[17] Wald, A. (1950), *Statistical Decision Functions*, New York: Wiley.

[18] Ziliak, S.T., and McCloskey, D.N. (2008), *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*, Ann Arbor: University of Michigan Press.